DOCUMENT RESUME

ED 464 956 TM 033 887

AUTHOR Jensen, Jacquelyn W.; Winitzky, Nancy
TITLE Exploring Preservice Teacher Thinking: A Comparison of Five Measures.
PUB DATE 2002-02-00
NOTE 28p.; Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education (54th, New York, NY, February 23-26, 2002).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Concept Mapping; Correlation; Higher Education; Interviews; *Measurement Techniques; *Preservice Teachers; *Teacher Attitudes; Teacher Education
IDENTIFIERS Self Report Measures; *Structure of Knowledge

ABSTRACT

Researchers use an array of measures to provide lenses and levels of magnification for exploring teacher candidates' thinking. This paper describes five measures in terms of their costs, benefits, and the kinds of data each yields. These measures include a demographic questionnaire, concept mapping, a learning questionnaire, stimulated recall interview, and a short answer assessment. The measures are illustrated through a study involving 16 female teacher candidates. Among the most significant findings was the degree of correlation among the various measures and the differing information between the stimulated recall and the learning questionnaire. Qualitative analyses of the concept maps reveal more cognitive change than do quantitative analyses. The moderate correlation among the three primary knowledge measures (concept maps, learning questionnaire, and short answer assessment) means that candidates' structural, declarative knowledge is related to candidates' explicit self-reports, which are in turn related to tests of the content. An examination of these measures can be used to make important methodological decisions in future investigations of candidates' learning. (Contains 2 figures and 13 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 956

Exploring Preservice Teacher Thinking:

A Comparison of Five Measures

Jacquelyn W. Jensen, Ph.D.
3934 West 7925 South
West Jordan, UT 84088
jwjensen62@yahoo.com

Nancy Winitzky, Ph.D.
Associate Professor
Department of Teaching and Learning
University of Utah
1705 E. Campus Center Dr., Rm 142
Salt Lake City, UT 84112
winitzky@ed.utah.edu

BEST COPY AVAILABLE

Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, February, 2002, New York, NY

TM033887

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. Jensen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract

What research tools do we presently have that allow us to know what candidates understand as they construct knowledge, and what instructional approaches they find most salient? Researchers use an array of measures that provide varying lenses and levels of magnification for exploring teacher candidates' thinking. A problem is that only a few comparisons of measures are available that give researchers a clear picture concerning the kind of data each provides.

The purpose of this paper is to describe five measures in terms of their costs, benefits, and the kinds of data each yields. These measures include a demographic questionnaire, concept mapping, a learning questionnaire, stimulated recall interview, and a short answer assessment. An examination of these measures can be used to make important methodological decisions in future investigations of candidates' learning.

Exploring Preservice Teacher Thinking:

A Comparison of Five Measures

In viewing learning to teach as skill learning (Winitzky & Kauchak, 1997), preservice teachers' declarative knowledge is important for the development of their procedural knowledge (Anderson, 1996). Teacher educators must, therefore, enable candidates to develop both kinds of knowledge, and pay special attention to helping them connect procedural and declarative knowledge so that they understand not only the how but the why of teaching (Gitlin, Barlow, Burbank, Kauchak, & Stevens, 1999; Jensen, 2001). To do so, teacher educators need to better understand how preservice teachers' knowledge develops in various learning environments.

What research tools do we presently have that allow us to know what candidates understand as they construct knowledge, and what instructional approaches they find most salient? Researchers use an array of measures that provide varying lenses and levels of magnification for exploring teacher candidates' thinking.[1] A problem is that only a few comparisons of measures are available that give researchers a clear picture concerning the kind of data each provides (e.g., Winitzky, Kauchak & Kelly, 1994). More information is needed concerning the interrelationships, strengths, and weaknesses of different measures.

The purpose of this paper is to describe five measures in terms of their costs, benefits, and the kinds of data each yields. These measures include a demographic questionnaire, concept mapping, a learning questionnaire, stimulated recall interview, and a short answer assessment. They were used to investigate the instructional strategies that were salient for teacher candidates and why these strategies influenced what they learned during a 2-

hour lesson on classroom management using small group theory. Specifically, the measures were used to answer four of the study's research questions that included: (1) What prior knowledge did teacher candidates have before instruction? (2) Which instructional strategies were salient for candidates? (3) Why were various strategies salient for candidates? and (4) What did teacher candidates learn from instruction? Note that this paper focuses more on the attributes of the measures used to answer these questions than on the study findings themselves. Such information advances our understanding because it can be used to make important methodological decisions in future investigations of candidates' learning. (See Jensen, 2001 for a description of the entire study).

Method

Sixteen female teacher candidates participated in the study. Candidates were enrolled in an integrated sequence of three methods and curriculum courses. They took these courses the semester before completing their student teaching in secondary schools. Participants attended a large, private university operated by a religious organization in the Rocky Mountains. A demographic questionnaire revealed that one-fourth of the sample were married; fifteen candidates were Anglo American and one was Hispanic; ten candidates were fluent in a language other than English. Their mean age was 24.

Dr. Adams, the instructor who participated in the study, was an Anglo American female in her late-50's. She was an associate professor and teacher educator in the School of Family Life, and had worked in the position since 1980. She team-taught the three courses with two faculty members from secondary education, and has taught these courses with one of them for the last decade. To determine what her goals were, two measures of her thinking

were obtained, an interview and a concept map.

On the first day of class, participants completed a demographic questionnaire, and constructed an unprompted and prompted concept map of their understanding of the lesson topic. Then 2 weeks into the semester, the instructor taught the lesson on classroom management. A video camera and field notes were the means used to gather naturalistic data about the instructional environment. At the end of the lesson, candidates constructed an unprompted and prompted concept map as was done on the first day of class. In addition to constructing the postmaps, candidates responded to a questionnaire. Finally, candidates compared their unprompted pre- and post-maps and wrote a paragraph indicating how their knowledge changed as a result of the instruction. While these postdata were being collected from candidates, the instructor also identified the concepts on her map that she felt were most and least emphasized during the instruction.

On two days following the instructional episode, four candidates selected at random participated in two stimulated recall sessions, one on each day. Finally, candidates completed a short answer assessment 6 weeks after the lesson.

Measures

Demographic questionnaire. Data from the demographic questionnaire provided an indication of candidates' background knowledge. The questionnaire asked participants their age, GPA, ACT score, gender, race, family income, prior teaching experience, prior education course work, languages spoken other than English, and the geographic location(s) where they had lived while growing up. Descriptive statistics (i.e., means and standard deviations) were used to analyze these data.

Concept mapping. The use of concept mapping assessed participants' structural knowledge or understanding of "the interrelationships among the important concepts" relative to classroom management (Goldsmith, Johnson & Acton, 1991, p. 88). An underlying assumption of concept mapping is that concepts are organized hierarchically in memory. The procedure for having participants construct concept maps was adapted from Eggen, Kauchak, Winitzky, Jensen, and Hadden (1997), Roehler, Duffy, Conley, Herrmann, Johnson, and Michelsen (1990) and Winitzky et al. (1994). To construct their concept maps, candidates watched a demonstration of the three-part procedure, and they received prototype maps from other content areas (see Figure 1). Candidates constructed an unprompted map first for the topic "classroom management." They brainstormed a list of terms related to the topic to activate prior knowledge, and organized them into a concept map depicting the relationships between concepts. After constructing the map, participants each described in writing the reason for organizing concepts as they did and specified relationships among them. Candidates then turned in their maps.

To construct a prompted map, candidates used as many terms as they liked from an alphabetized list of terms or target concepts generated from the instructor's map. We told them that they could integrate additional items not on the list. Again, candidates wrote a description to explain their reasons for organizing concepts as they did and to specify relationships among them. Then at the end of the classroom management lesson, candidates constructed an unprompted and prompted concept map as was done on the first day of class.

To analyze candidates' maps, we used scoring procedures from Roehler et al. (1990) with one modification suggested by Winitzky et al. (1994). The modification, a means of strengthening validity, consisted of dropping the

Concepts Per Chunk subscore. Further, we did not compute Coherence due to the correlation of Chunk Coherence and Sequence Coherence with the Roehler et al. Overall score (Winitzky et al., 1994). Thus, we scored the maps for: (a) the number of individual concepts, (b) the number of chunks, i.e., groups of superordinate concepts with two or more subordinate concepts, and (c) a hierarchical structure score, i.e., the sum of the number of horizontal chunks at the widest level and the number of vertical levels (Eggen et al., 1997).

To increase scoring reliability, we generated the following decision rules: (a) a chunk (i.e., a group of superordinate concepts with two or more subordinate concepts) can radiate up, down, or to the side, (b) to increase scoring reliability, a phrase or words listed together will count as one concept, with the exception of prompted terms, (c) a concept listed twice can be counted twice if the candidate has not linked it to the same concept twice, (d) a string of concepts that are not hierarchically related will be analyzed as if they fan out from the concept, (e) the widest point (used in calculating hierarchical structure) is the level where the most chunks reside, not the line having the most concepts, and (f) if a chunk has concepts that radiate both up and down, assume that they radiate down when calculating chunks at the widest point.

Prior to conducting a reliability check with a random sample of maps, one of us examined each map on four separate occasions. To check for scoring reliability, a random sample of four maps were scored, and interrater agreement was 100%. Then the two of us individually scored a random sample of maps followed by negotiation. Discussion occurred until raters came to complete agreement.

After all maps were scored, we examined them both quantitatively for

structural complexity and qualitatively for content. Quantitatively, we made comparisons between pre- and post-maps for changes in the number of unprompted concepts, prompted concepts, chunks, and for changes in hierarchical structure. Qualitatively, we made comparisons between pre and post maps for: (a) the presence of target concepts (i.e., those from the instructor's map), (b) the use of target concepts under both post unprompted and prompted conditions, (c) the prominence of specific target concepts, and (d) the durability of candidates' non-target concepts from pre to post mapping (Eggen et al., 1997).

In studies by Eggen et al. (1997), Michelsen (1987), Naveh-Benjamin and Lin (1994), Roehler et al. (1990), and Winitzky (1992), validity in the form of criterion-related evidence is apparent suggesting that "test scores are systematically related to one or more outcome criteria" (Standards for educational and psychological testing, 1985, p. 11). Further, the findings from Winitzky et al. (1994) and the series of Roehler et al. studies provide some evidence for construct validity. Evidence for several kinds of reliability using concept mapping was apparent in the Eggen et al. (1997), Roehler (1990), and Winitzky et al. (1994) studies.

Learning Questionnaire. In addition to constructing the postmaps, candidates responded in writing to a questionnaire that contained three questions. These asked candidates to identify (a) the most important concept(s) they had learned, (b) the features of instruction that helped them to learn each concept, and (c) why these aspects of instruction helped them to learn the concept(s).

To analyze candidates' responses on the first question where candidates identified the most important concept(s) they learned during the lesson, one

of us coded each candidate's response twice to ensure reliability, and then met with the instructor who independently coded the data. We discussed the few instances in which we had not applied the same code(s) until reaching full agreement. Codes for these data included the 28 target concepts from the instructor's map.

Codes for analyzing candidates' responses to the second question concerning those instructional features they found salient were generated inductively. Again, one of us coded candidates' responses twice to ensure reliability. We then discussed the codes and decided to divide one of them into two codes because candidates' responses suggested they had learned from at least one of two strategies that have varying assumptions. One of us then recoded candidates' responses on two more occasions before we conducted a reliability check. Interrater agreement was 100%.

We deductively categorized candidates' responses on the third question concerning why the instruction helped them to learn using theory-driven codes. Nevertheless, we also remained open to the possibility that other codes could be derived if we determined that a category of data was not being represented by those codes previously identified. Most of the 12 codes were based on information processing and schema theory. In addition, one code was based on the motivation literature, and another was based on the Posner et al. (1982) model of conceptual change learning.

Stimulated recall interview. Two days following the lesson, 4 candidates selected at random participated in two stimulated recall sessions, one on each day. We had candidates participate in two sessions because of the time needed to view the 2-hour videotape. A 2-hour session was required for each 1-hour lesson segment. To prevent fatigue, candidates took a five minute break after

the first hour of each session. We conducted the stimulated recall sessions on the days immediately following the lesson because we wanted to diminish memory decay, an important consideration when having candidates reflect on their learning. Winitzky and Kauchak (1995) hypothesized that over time, "the declarative knowledge acquired in university course-work becomes tacit knowledge subsumed into productions and unavailable to conscious, explicit retrieval" (p.223).

Three candidates participated in the sessions with the first author and 1 candidate participated in them with a trained research assistant. We viewed a videotape of the instruction with each participant using a videocassette recorder (VCR) and a television monitor. Each candidate was instructed to stop the videotape in places where the instruction best aided her learning, explain why the instruction was helpful and identify what was learned. The procedure of having candidates stop the tape is new to this study. Researchers generally identify the critical incidents on which they want candidates to reflect.

We audiotaped the stimulated recall sessions. When each candidate reached to stop the VCR, we wrote down the time shown at the bottom of the screen and the key words used by candidates in responding to the three questions. These procedures helped to facilitate transcription and analysis. Further, we probed carefully every time the candidate stopped the tape to ensure that she had answered all three questions.

To analyze candidates' responses concerning the instructional features that helped them to learn, one of us condensed onto matrices the portion of each of the 4 candidates' interview responses that adequately answered this question, and generated codes inductively. After coding each candidate's

responses on three separate occasions, we met to conduct a reliability check. We negotiated the few instances where we had not applied the same codes until reaching full agreement.

Similarly, candidates' responses during stimulated recall sessions concerning why the instruction helped them to learn were also condensed onto matrices, but then categorized deductively using theory-driven codes. As with the like question on the questionnaire, we based our codes primarily on information processing and schema theory. Again, during the application of codes, we remained open to the possibility that we could derive other codes if we determined that a category of data was not being represented by those that we had previously identified. A few additional codes, therefore, correspond with the literature on metacognition, conceptual change learning (Posner et al. 1982), and motivation.

To analyze candidates' stimulated recall responses concerning what they learned, the 28 target concepts from the instructor's map served as codes. In addition, we used the code "incidental learning" when a response yielded no evidence that the candidate had learned a target concept(s). After one of us had coded each interview twice to ensure reliability, the instructor coded the data. We discussed all instances in which we had not applied the same code(s) until reaching full agreement.

Further, at the end of the first stimulated recall session, we asked candidates three questions to assess their beliefs about teaching and learning generally. To generate these questions, we examined studies that explored candidates' conceptions of teaching and learning. We then asked two colleagues to help us select and refine several questions. Those we all agreed upon included: (a) Imagine you are teaching in your ideal setting.

Describe a typical class period. (b) How do people learn? and, (c) What are the most important things that teachers do to help students learn? To analyze these data, the first author generated codes for candidates' responses based on learning theories. We then met together to read each candidate's responses, examine the use of codes, and brainstorm additional codes to account for other aspects for teaching.

Short answer assessment. All 16 candidates completed a short answer assessment 6 weeks after the instructional episode. The assessment provided a distal measure of the degree of congruence between the instructor's goals and candidates' understandings. Dr. Adams helped to develop the essay questions and rating scales for analyzing candidates' responses. We used guidelines specified in Stiggins (1997) for scoring open response formats.

The process used to develop essay questions began with Dr. Adams writing four objectives and a range of questions for each. She then selected a representative sample of one question for each objective. To score the questions, we modified a scale provided as an example by Stiggins (1997, p. 169) so that each description placed more emphasis on the completeness of candidates' responses (see Appendix). To aid in the scoring of candidates' responses, Dr. Adams wrote or outlined a quality response for each question. Each candidate's response to the first question was scored before beginning to score responses to the second question. Candidates' responses to each question were scored in this manner before moving on to subsequent responses. After scoring candidates' responses on two separate occasions, one of us met with the instructor. With the ideas and decision rules contained in several memos clearly in mind, Dr. Adams independently scored each question. We discussed the few discrepancies in our scoring until reaching full agreement.

Findings

The demographic questionnaire provided general information concerning candidates' life experiences and was an indication of their prior knowledge (see Table 1). For example, candidates had taken 1.69 education courses prior to the study. This average suggests that candidates' knowledge concerning teaching was limited prior to participating in the study. Concept maps from the pre-test, in contrast, provided information that was more specific in terms of candidates' knowledge of classroom management (see Figure 2).

Information provided by the quantitative and qualitative analyses of the concept maps is of particular interest. The most noticeable difference in the findings was that the quantitative analysis of concept maps suggested little or no change in candidates' knowledge structures. Previous studies equate increases in hierarchical structure, concepts and chunks with increased cognitive organization. Yet, rather than smoothing out the differences in candidates' learning, the qualitative analysis of these data accentuated the changes that occurred among target and idiosyncratic concepts from pre- to post-mapping. That is, candidates' use of target concepts increased (unprompted=59.12%; prompted=20.93%) and their use of idiosyncratic concepts decreased (unprompted=29.52%; prompted=82.41%).

Candidates were asked essentially the same questions within both the learning questionnaire and the stimulated recall interview. Nevertheless, the stimulated recall interview data allowed for a more fine-grained examination than did data from the questionnaire. The stimulated recall interview enabled candidates to stop the videotaped lesson in places where the instruction helped them to learn something and respond to the questions mentioned previously. Doing so provided a large amount of data.

The learning questionnaire, in contrast, asked candidates to reflect on their learning immediately following the lesson. The nature of the instrument enabled candidates to give short responses. The responses provide, therefore, few specific details about candidates' learning, possibly because of the limitations of memory given the task and/or the small space provided for each candidate's written response. That is, the instrument may not have enabled candidates to recall detailed information. The stimulated recall interviews allowed minute-to-minute, detailed reflections whereas the questionnaire allowed for brief, general reflections of learning. A major difference in the two measures seemed to be the level of magnification.

In comparing concept mapping with the learning questionnaire and the stimulated recall interview, concept mapping accessed a broader sweep of the concepts candidates knew. The questionnaire focused candidates on identifying "the most important concepts" they had learned rather than focusing them on all of the concepts they understood. Further, the short answer assessment contained a small representative sample of a larger number of questions. Of the measures, the stimulated recall interviews provided the most fine-grained lens for examining candidates' thinking.

We wondered how candidates' performance on each measure compared with their performance on other measures, i.e., unprompted mapping, prompted mapping, learning questionnaire, short answer assessment. (Recall that stimulated recall interviews were only conducted with 4 candidates.) That is, did students score or respond consistently, either high or low, across measures? A qualitative analysis revealed that candidates did indeed perform similarly on all four measures.

We also looked at the relationship of these measures using the Pearson

product-moment correlation coefficient. The correlations are fairly high, suggesting that the measures are related. (Correlations ranged from .46 to .61). Since the sizes of the relationships are not in the .80 or .90 range, however, they also suggest that each instrument measures something unique. For example, an examination of target concept frequency counts from pre- to post-prompted mapping suggests that the concepts most salient for candidates were the two small group methods. Candidates' responses on the learning questionnaire, in contrast, suggests that of the lesson content the most salient information was related to why they should use small groups of students to manage the classroom. In sum, concept mapping accessed candidates' focus on specific skills whereas the learning questionnaire accessed their focus on reasons for using them.

Conclusions

Researchers use several measures as they examine candidates' thinking because of the difficulty in ascertaining what someone knows. Too few descriptions of the differences in the kinds of data yielded by various measures are available in the existing literature. This paper documents differences in five measures including a demographic questionnaire, concept mapping, a learning questionnaire, stimulated recall interview, and a short answer assessment.

The most significant findings are:

- the degree of correlation between the various measures,
- the differing type of information from the stimulated recall versus the learning questionnaire,
- qualitative analyses of the concept maps reveal more cognitive change than the quantitative analyses.

There was a moderate correlation between the three primary knowledge measures - the concept maps, learning questionnaire, and short answer assessment. This means that candidates' structural, declarative knowledge, as measured by the concept map, is related to candidates' explicit self-reports, which are in turn related to tests of the content. This confirms previous research in yet another, novel population, suggesting robust generalizability (Naveh-Benjamin & Lin, 1994; Winitzky, et al., 1994). This is also very helpful in ensuring validity through triangulation of multiple measures.

Since the measures are related but not identical, researchers have great flexibility in designing studies exploring the relationship between learning environments and teacher candidate learning. With large Ns and experimental designs, whose purposes are to assess the power of particular instructional strategies, we recommend concept mapping (quantitatively analyzed), coupled with content tests, because of the reduced labor costs associated with analysis. If the focus, however, is on the internal processes of teacher candidates as they strive to construct meaning from a particular learning environment, then investigators are advised to utilize concept maps in a qualitative fashion coupled with the learning questionnaire. These measures provide more detail but are more time consuming to analyze.

Parallels can be found in comparing the stimulated recall measure with the learning questionnaire. Stimulated recall provided far more detail on the cognitive processes of the participants, but necessitated great investments of time and energy from both researchers and candidates. Again, case study research where the goal is understanding indicates stimulated recall, while large-scale, descriptive or experimental research indicates the shorter questionnaire. As in this study, a nested design, which provides both the

broad strokes and the fine detail, can utilize both to good result.

A significant finding is the difference between the results obtained through quantitative versus qualitative analysis of the concept maps. Recall that the numeric scores did not reveal substantial change in knowledge from pre to post, while the qualitative analysis showed credible evidence of knowledge growth in the desired direction. While this is rare, it is not unheard of (see Winitzky & Kauchak, 1995, for example). How do we make sense of this, and what does it mean for researchers?

There are at least a couple of reasons why the concept map scores may have showed little change. It may be too coarse a measure to demonstrate change in a small group (N=16) over a short period of time (two weeks). Study duration in most previous research ranged from one semester to one year. Some scholars have speculated that in the early phases of learning, teacher candidates are struggling to integrate new information into existing knowledge structures, and that that results in both turbulence (high turnover of concepts over time) and lack of observable change in numeric scores (Winitzky & Kauchak, 1997). That is, their scores are low initially because they don't know much, and their scores are low later because they haven't completed the knowledge integration process. If this is the case, researchers are strongly advised to include qualitative along with quantitative analyses in order to ensure that all growth is observed and recognized.

An urgent need in teacher education is to strengthen the power of our programs and thereby the skills, knowledge, and wisdom of our graduates. Without a solid understanding of the connection between the learning environment and candidate learning, this will be impossible. This study has shown how cognitive measures like concept mapping, stimulated recall, and the

learning questionnaire can be used most efficiently to explore that critical question. We encourage other researchers to further define, refine, and build upon this work to develop more powerful ways to understand teacher candidates' learning and learning processes.

References

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. American Psychologist, 51, 355-365.

Eggen, P., Kauchak, D., Winitzky, N., Jensen, J., & Hadden, J. (1997, April). An exploratory study of conceptual change in an introductory educational psychology course. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Gitlin, A., Barlow, L., Burbank, M., Kauchak, D., & Stevens, T. (1999). Preservice teachers' thinking on research: Implications for inquiry oriented teacher education. Teaching and Teacher Education, 15, 753-769.

Goldsmith, T., Johnson, P., & Acton, W. (1991). Assessing structural knowledge. Journal of Educational Psychology, 83, 88-96.

Jensen, J. W. (2001). Constructivism and supporting environments: Exploring instructional salience in preservice teacher education. Unpublished doctoral dissertation, University of Utah, Salt Lake City.

Naveh-Benjamin, M., & Lin, Y. (1994). Measuring and improving students' disciplinary knowledge structures. In P. R. Pintrich, D. R. Brown, & C. E. Weinstein (Eds.), Student motivation, cognition, and learning: Essays in honor of Wilbert J. McKeachie (pp. 51-78). Hillsdale, NJ: Lawrence Erlbaum.

Roehler, L. R., Duffy, G. G., Conley, M., Herrmann, B. A., Johnson, J., & Michelsen, S. (1990). Teachers' knowledge structures: Documenting their development and their relationship to instruction (Research Series No. 192). East Lansing: Michigan State University, Institute for Research on Teaching, College of Education.

Standards for educational and psychological testing. (1985). Washington,

DC: American Psychological Association.

Stiggins, R. J. (1997). Student-centered classroom assessment (2nd ed.). Upper Saddle River, NJ: Merrill.

Winitzky, N. E. (1992). Structure and process in thinking about classroom management: An exploratory study of prospective teachers. Teaching and Teacher Education, 8(1), 1-14.

Winitzky, N. & Kauchak, D. (1997). Constructivism in teacher education: Applying cognitive theory to teacher learning. In V. Richardson (Ed.), Constructivist teacher education: Building new understandings (pp. 59-83). Washington, DC: Falmer.

Winitzky, N., & Kauchak, D. (1995). Learning to teach: Knowledge development in classroom management. Teaching and Teacher Education, 11(33), 215-227.

Winitzky, N., Kauchak, D., & Kelly, M. (1994). Measuring teachers' structural knowledge. Teaching and Teacher Education, 10(2), 125-139.

Footnotes

[1]Table 2 in Jensen (May, 2001) catalogues the sources of data used in 47 studies that explore the degree of change in candidates' conceptions and practices during teacher education. The most common data sources include interview, observation, videotape, concept mapping, lesson plans, and questionnaire.

Table 1

Selected Background Statistics for the Sample

Variable	Mean	SD	Comment
Age	24.06	3.86	Range = 21-34
GPA	3.4	.30	Obtained by self-report
ACT	26.63	3.77	Range = 21-34
Annual Income	$11,517.33		Range = $3,960.-$37,000.

Appendix

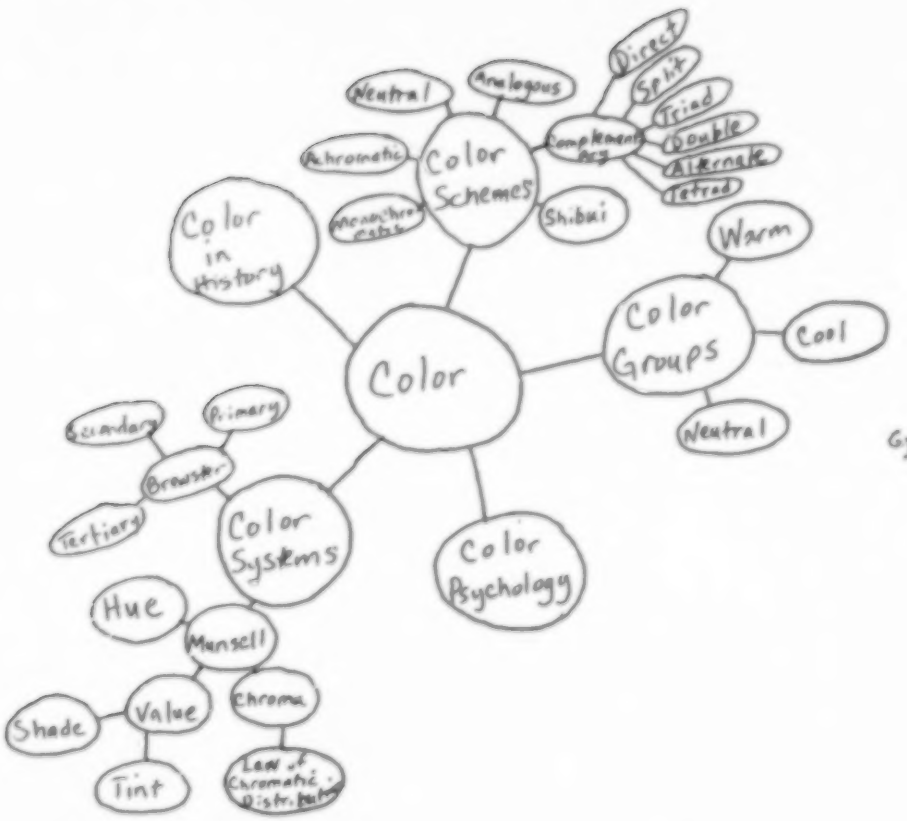
Note: If a response falls in between two given scores, it is perfectly acceptable to assign the response with 2 or 4 points.

5 The response is clear, focused and accurate. Relevant points are made (in terms of the kinds of reasoning sought by the exercise) with good support (derived from the content to be used, again as spelled out in the exercise). Good connections are drawn and important insights are evident. The response is complete.

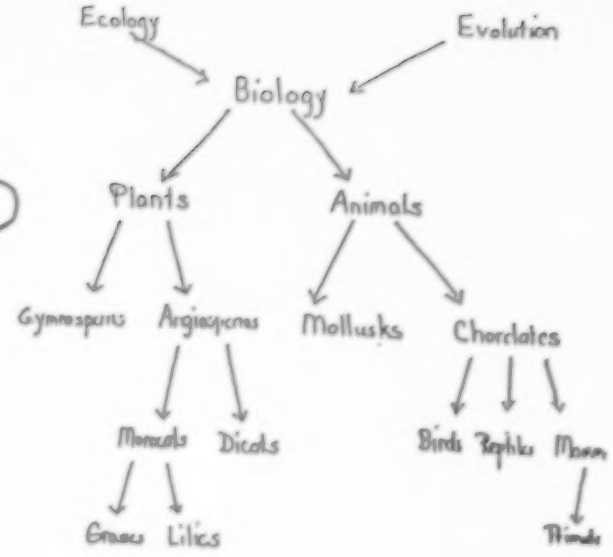
3 The answer is clear and somewhat focused, but not compelling. Support of points made is limited. Connections are fuzzy, leading to few important insights. The response is incomplete.

1 The response either misses the point, contains inaccurate information, or otherwise demonstrates a lack of mastery of the material. Points are unclear, support is missing, and/or no insights are included.

Adapted from an example in Stiggins, R. J. (1997). Student-centered classroom assessment, p. 169. Upper Saddle River, NJ: Merrill.



25



26

Figure 1. Examples of concept maps.

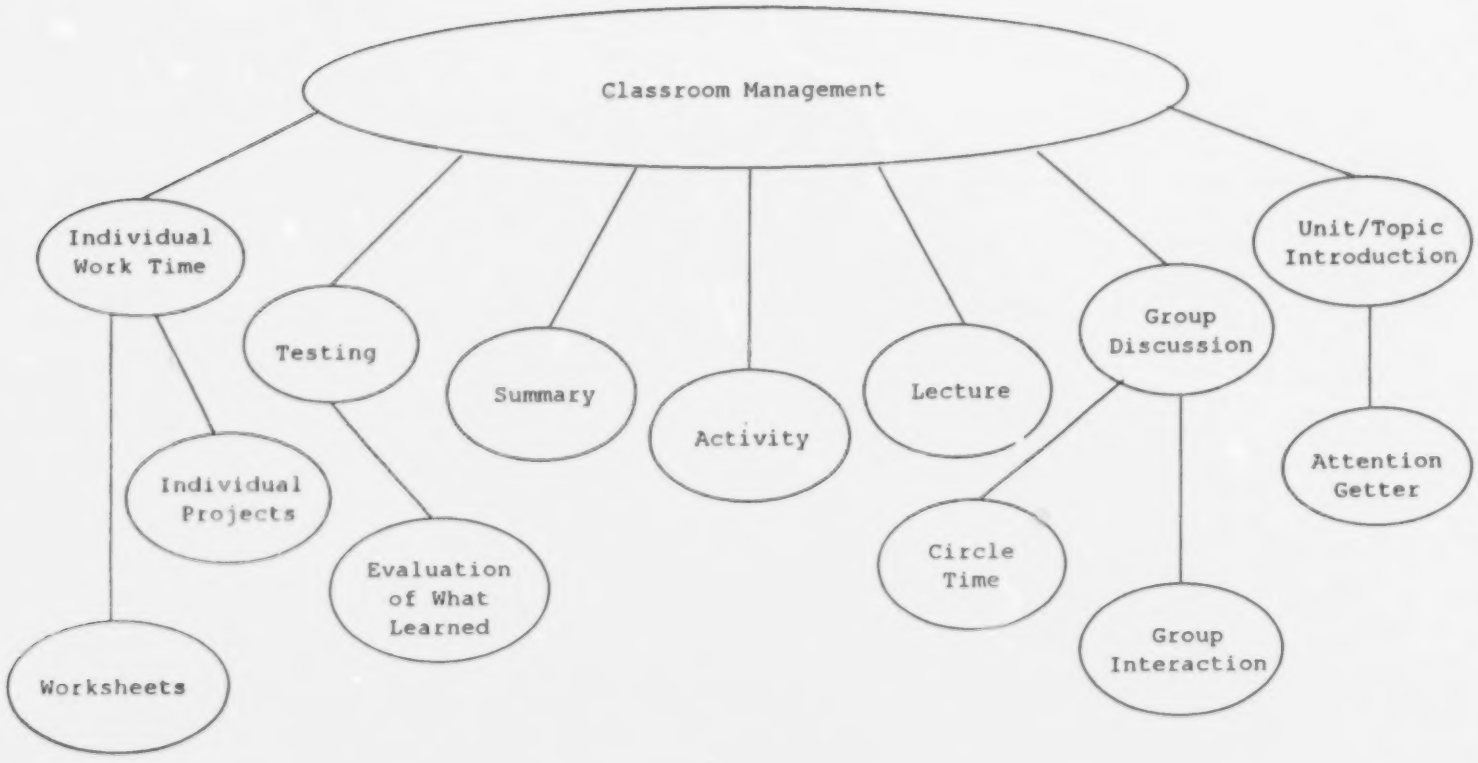


Figure 2. A candidate's preunprompted concept map.